for American workers. Citizens for HDTV (at 12) reiterates these concerns, noting that the European Commission has already issued a binding directive that a single digital transmission standard (DVB) will be used in cable and DBS, and a similar directive is expected soon for terrestrial broadcasts.[67]

General Instrument (at 8-11) summarizes the history of the international trade aspects of the HDTV proceeding, and urges the Commission to help ensure that the standard is finalized expeditiously, promoted first throughout North America and then in South America and Asia, and supported in specific cases where DVB, although inferior to the ATSC Standard, is making crucial inroads.

Sony (at 11-12) argues powerfully that a critical issue of American leadership is at stake, and that a mandated standard is essential. Dolby (at 4) observes that a small delay can be explained, but failure to mandate a standard could soon cripple efforts to export the ATSC standard. Universal Studios (at 2) notes that by incorporating the standard in its rules, the FCC will lay the foundation for enhancing the position of U.S. program producers, and MPAA (at 8) argues that the standard will facilitate international program exchange.[68]

All of these comments demonstrate conclusively that the rapid adoption of the ATSC DTV Standard will promote international trade and improve our nation's international

---

[67]*See also,* Reply Comments of the North American National Broadcasters Association, August 9, 1996, saying that further delay in establishing a standard will only hurt North America's ability to participate optimally in the benefits of leadership associated with digital transition.

[68]A few parties to this proceeding have embraced the mistaken notion that adopting the ATSC Standard would somehow help our foreign trading partners at the expense of Americans. For example, the American Homeowners Foundation (at 2) says that more jobs for American homeowners will be created by policies that increase the U.S. demand for computers than will be created by policies that increase demand for TVs, since more demand for the latter mostly creates jobs for workers of our trading partners. This type of strained, convoluted comment arises out of erroneous and misleading statements made by some members of the computer industry with respect to interoperability issues. The Commission's goal is not and should not be to handicap one industry against another, nor would any of the interoperability issues in this proceeding have that effect, contrary to the assertions of some. The Commission's primary purpose in this proceeding is to oversee the upgrading of free over-the-air television, and to ensure that a competitive marketplace operates to give consumers cost-effective options for accessing that service, including the option to buy a low-cost, basic TV, a more expensive top-of-the-line TV, or a combined PC/TV product.

competitiveness, spurring economic growth and the creation and preservation of high-paying jobs for Americans.

## IX. Conclusion

As these reply comments have amply demonstrated, none of arguments against adopting a complete standard, and none of the complaints raised against the ATSC DTV Standard in particular, and certainly not the CICATS counterproposal, nor anything else in the voluminous comments on the NPRM provides a sound basis for changing the Commission's tentative decision to adopt the ATSC DTV Standard as the single standard for use by digital broadcast television licensees. In fact, this thorough analysis of the comments demonstrates conclusively that the Commission should fully embrace the recommendation of its Advisory Committee and adopt the ATSC DTV Standard in its entirety. By so doing, the Commission will unleash a flurry of investment within the involved industries that will support a rapid implementation of digital broadcast television, quickly bringing the fruits of this beneficial new technology to the American public and beyond.

Accordingly, the Commission should adopt the ATSC DTV Standard for terrestrial broadcast transmission as rapidly as possible.

Respectfully submitted,

Robert K. Graves

Robert K. Graves
Chairman

Mark Richer

Mark S. Richer
Executive Director

Advanced Television Systems Committee

1750 K Street, NW, Suite 800
Washington, DC 20006
(202) 828-3130
(202) 828-3131 (fax)
http://www.atsc.org

August 12, 1996

# APPENDIX A: CICATS MPEG COMPLIANCE ISSUES

## A. Fundamental violations of the MPEG standard (and some discussion of consequences):

1) Temporal level enhancements, which are proposed by CICATS to be sent in a pure B-frame bit stream, are incompatible with the MPEG standard. It is also not clear how the CICATS system identifies where the temporal enhancement B-frames belong in the image sequence

2) Spatial enhancements, as proposed by CICATS, are incompatible with the MPEG standard. As we understand the CICATS system, it requires more expensive interpolation filters. CICATS' interframe coding of enhancement information requires more precision and dynamic range than MPEG spatial scalability.

3) CICATS would prohibit interlace scan, frame rates of 23.976, 29.97, 30, 59.94, and 60 Hz, and non-square pixel formats at 4:3 and 16:9, all of which exist in MPEG and are embodied in currently available video products purchased by consumers. *CICATS receivers and converters would be unable to receive every known bit of digital TV that is currently transmitted in the U.S. via satellite, cable, MMDS, DVD or telephone company video delivery systems.*

4) B-frames are disallowed in the CICATS base level. This necessarily limits encoder flexibility in facilitating reduced-cost decoding of the base level. Prohibition of B-frames also limits options for support of necessary consumer VCR functions, including extraction of subsets of the video data for trick play functions. The restriction should be particularly onerous to computer CPU-based decoders, since it removes a number of options for graceful decoder degradation. In addition, it is well-accepted that the inclusion of B-frames provides higher quality images for a given bit rate. This is especially true for more critical video sequences.[1]

5) No maximum bit rates are specified. Clearly specified maximum bit rates are needed to ensure interoperability.

6) CICATS requires that film color primaries must be supported, in contrast to MPEG's support of video color primaries. Since the video transmitted is not intended for film transfer, this would seem to be an unnecessary burden on receivers.

7) CICATS appears to require display functions (e.g., overlay capability) that will add cost.

8) The CICATS system is poorly specified, with ambiguities and missing information. This stands in sharp contrast to the formal specifications of the MPEG and ATSC Standards. The MPEG standard (and its ATSC derivative) also had the benefit of dozens of the world's experts on video compression

[1] "Performance Evaluation of MPEG-2 Video Coding for HDTV," Daniel Lauzon, Andre Vincent and Limin Wang, Communications Research Centre, IEEE Transactions on Broadcasting, June, 1996.

checking each others' work and documentation, including interoperability experiments and objective quality experiments.

**B. Violations of any currently defined MPEG profile or level:**

1) The only frame rates allowed in any currently defined MPEG profile or level are 24, 25, 30, 50, and 60 Hz (and the 23.976, 29.97, and 59.94 Hz "TV-rate" variations). Thus, the 72 Hz and even the 36 Hz CICATS base level formats are not compliant with any existing MPEG-2 profile or level.

2) CICATS allows up to 2048 horizontal pixels, whereas the MPEG-allowed maximum is 1920.

3) The CICATS base level specifies a 1 Mbit channel buffer. This is too small to permit high quality video decoding. MPEG-2 MP@ML specifies a 1.835 Mbit channel buffer, which is at least 4.4 bits/pixel, but CICATS only allows 1.9 bits/pixel. We note that this issue of channel buffer size can easily be overlooked in simulation. Had DemoGraFX implemented its system in real-time hardware, the situation would have become apparent.

## C. Violations of MPEG Main Profile and Main Level:

1) MPEG MP@ML specifies a maximum pixel rate of 10.368 Mpixels/second. CICATS allows up to 18.88 Mpixels/second in the base level. CICATS claims that there are MPEG-2 MP@ML decoders that "will go as fast as 20 Mpixels/second or more". We have been unable to find any.

2) The CICATS base level can have a horizontal pixel count as high 1024, but the corresponding MPEG-2 MP@ML limit is 720.

**Mark Schubin**
**Technological Consultant**
40 West 72nd Street #43
New York, NY 10023-4104 USA
Phone: +1-212-315-2850
Fax: +1-212-870-4520
E-mail: mschubin@mcimail.com

8 August 1996

Robert K. Graves, Chairman
Advanced Television System Committee
1750 K Street, N.W.
Washington, D.C. 20006

Dear Mr. Graves:

Enclosed please find the latest version of my paper "Searching for the Perfect Aspect Ratio." It is being published in this month's issue of the *SMPTE Journal*. I've been told that both the film-background and video-background reviewers considered it an excellent tutorial. You may be aware that I first presented the paper at World Media Expo in New Orleans last year as "The History of the Perfect Aspect Ratio." I have since lectured and published more on the subject.

Based on my extensive research, I have concluded that no aspect ratio is preferable to any other from an aesthetic, psychological, or physiological standpoint. I have, however, found misstatements on both sides of the debate. In the 16:9 camp, these misstatements are primarily related to the concepts of equal-area shoot & protect production and polyscreen/multiple-picture-in-picture display. There is nothing sacred about 16:9 for either of those. There is also some hedging on the significance of 16:9's relationship to the international digital video standard Recommendation 601; 16:9 works well with that standard in the horizontal domain but not so well in the vertical.

That's about it for 16:9 being pushed too hard. The 2:1 camp, however, seems to be way off in just about everything they say. First and foremost, 2:1 absolutely does ***NOT*** solve the letterbox problem for the vast library of 2.4:1 and 2.2:1 films. Any reduction in the problem for those films is at the sacrifice of the display of the even larger libraries of films shot at 1.85:1 or less. It is also not true that wider screen movies earn higher revenues than do narrower ones. At the time of my research last year, the opposite was actually true, by a wide margin, based on *Variety*'s lists of the top-grossing movies of the year and of all-time. It's not true that the SMPE long ago approved a 2:1 widescreen aspect ratio; as my paper shows, they approved 1.8:1, and they showed it on a 16:9 screen. It's also untrue that cinematographers have always favored 2:1; my paper shows that they actively opposed it in the past. As recently as 1995, the great cinematographer Walter Lassally made a plea for 1.75:1 (essentially 16:9)

The 2:1 camp does have a legitimate, but trivial, point about the value of expressing an aspect ratio as 16:9 instead of as 1.78:1. Pioneer is promoting its 1.5:1 widescreen TV sets as having an aspect ratio of 16:10.7. There can be no reason for such an obscure choice of numbers other than the implication that it's even better than 16:9. I do not believe that's why anyone selected 16:9 over 2:1, however, and the use of integer values is explained in my paper.

There is much to be said for sticking with the existing television aspect ratio of 4:3, there's something to be said for an aspect ratio of 2.4:1 (it would essentially eliminate letterbox and allow theatrical-style masking to be used on video displays), and, of course, there's a great deal to be said for 16:9, not least of which is its international standardization. There is essentially only one thing going for 2:1. If any relationship to Recommendation 601 is ignored, 2:1 fits precisely into two-megapixel memories.

Please let me know if you need any more information. This is a subject near and dear to my heart, and I hate seeing all the unresearched misinformation spread about it.

Sincerely,

# Searching for the Perfect Aspect Ratio

Mark Schubin
Consultant
New York, New York

## Abstract

A debate is currently taking place over the appropriate aspect ratio for advanced television displays. Any selected aspect ratio is inherently incompatible with any other and will require the use of some form of accommodation technique. The derivation of the 16:9 (1.78:1) aspect ratio from accommodation techniques and display modes is explained, as is the relationship between aspect ratio and display memory. Research into the history of aspect ratios indicates that the 1.78:1 aspect ratio was adopted by the standards committee of the Society of Motion-Picture Engineers in 1930. It also indicates that the factors that may initially have led to widescreen motion-picture systems may no longer be applicable. The research for this paper found no clear indication of a preference for any particular aspect ratio for moving images nor any physiological reason to favor one over another. The research did show that cinematographers have not always favored the same aspect ratio.

A 1988 paper entitled "Another Method of Aspect-Ratio Conversion For Use In Receiver-Compatible EDTV Systems" begins. "Two systems with different aspect ratios are inherently incompatible"[1] [EDTV is extended-definition television]. The statement bears looking into.

For the purposes of this paper, *aspect ratio* will be defined as the ratio of an image's width to its height. Ever since there have been rectangular images, there have been aspect ratios (and it may be argued that even elliptical images have aspect ratios).

We are surrounded daily by multiple aspect ratios not seeming to cause any incompatibility problems. Images in newspapers and magazines have a variety of aspect ratios both greater and less than one; the same is true of paintings and photographs. Even some computer display screens may be rotated from a horizontal aspect ratio ("landscape") to a vertical one ("portrait"). When theatrical motion-picture and television screens are considered together for the purpose of displaying the same imagery, however, the inherent incompatibility becomes more clear.

That incompatibility became most noticeable in 1961, when the 1953 CinemaScope movie *How to Marry a Millionaire* was broadcast on the NBC television network.[2] Intended to be seen at an aspect ratio of 2.55:1 (and with image composition intentionally filling the wide frame), the movie was truncated to television's 4:3 (1.33:1). Almost immediately, technical publications began to carry information about how best to deal with the "conversion" of one aspect ratio to another.[3]

## Aspect-Ratio Accommodation

In fact, imagery is not "converted" from one aspect ratio to another; one aspect ratio is merely accommodated by another, almost invariably with some degradation of the imagery involved.[4] There are only three basic methods of accommodating existing material shot in a fixed aspect ratio on a display of a different fixed aspect ratio, though the techniques may be combined. These three basic techniques are shown in Figure 1.

Figure 1-A shows the truncation method, a variant of which is sometimes referred to as "pan & scan." When going from a wider aspect ratio to a narrower one in this method, the heights of the two images are matched, and any excess width in the wider image is removed from the display. The position of the displayed rectangle in the pan & scan mode may vary either by gradual "panning" (or "tilting," in the case of accommodation of a narrower aspect ratio) or by rapid repositioning ("cutting") between frames.

Figure 1-B shows the shrinking method, referred to as "letterbox" due to the shape of the shrunken image window when a wider aspect ratio is being accommodated on a narrower display. The black bands need not be evenly spaced. It is often the case that the lower band (when a wider aspect ratio is being accommodated) is made larger for the purpose of carrying subtitles, and, as will be discussed later in this paper, when a narrower ratio is being accommodated, the elimination of one of the side bands offers the possibility of stacking additional images in the other side band, a technique that has been referred to as multiple picture-in-picture (MPIP).

Figure 1-C shows the distortion method, whereby the linearity of the geometry of the image is changed to squeeze it into a different display shape. In a recent variation on this technique, a non-linear distortion is used, affecting the edges of the image more than the center (e.g., in JVC's NV-55BH6 and NV-55BX6 consumer widescreen projection receivers).

As can be seen from the two rows of Figure 1, the same basic methods apply whether the original image is wider than the display or narrower. In fact, the same techniques apply whether the two aspect ratios are both film, both video, or one of each. From 1961 to date, however, generally only the techniques of the upper row have been seen as widescreen movies have been shown on narrower video screens in homes, aircraft, or other venues. Unless otherwise specified, the word *widescreen*, for the purposes of this paper, will be used as defined by the British Kinematograph Sound and Television Society (BKSTS), "in general, pictures presented with an aspect ratio greater than 1.4:1."[5]

All of the accommodation techniques of Figure 1 are problematic. Sometimes aspect ratio accommodation is demonstrated with so-called "neutral" imagery: pictures that appear no less desirable when cropped. Motion pictures and television shows are not shot to be neutral, however. The truncation technique clearly causes portions of the image to be lost, and the variants associated with pan & scan introduce motion or cutting never intended in the original.

The distortion technique clearly changes the shape of not only the image but also people and objects contained within it. An informal survey conducted in association with the research for this paper found that distortion in the range of two to six percent may be considered acceptable, but that is much less than the amount needed to accommodate a typical widescreen movie on a conventional television display or vice versa.

The shrinking technique (letterbox) preserves the original image composition but reduces the visual angle available to the viewer and often resolution as well. Detail that is just perceptible in an image when it is viewed at a particular display resolution will be lost if the same image is shrunk on the same display.

In most cases when the viewing screen is video-based, this shrinking results in noticeably empty portions of the display device, a condition that has been considered objectionable to audiences by some television programmers.[6] One television set manufacturer has introduced a widescreen model (JVC AV-36W1) with a mechanical masking system that covers the unused portions of the display much as drapes mask unused portions of some motion-picture theater screens, possibly resulting in the reduction or elimination of such objections.[7]

A potentially more serious problem related to the shrinking technique is differential phosphor luminance decay, a reduction in the light output of the cathode-ray tube phosphors in the active picture section relative to that in the blank section, often affecting blue phosphors more than red or green.[8] As a result, when the full display area is viewed, the shrunken image area can become visible as a stripe somewhat yellower than the rest of the display. The effect is greater in projection displays than in direct-view displays due to the higher beam currents of the former.

It has been suggested that the differential phosphor luminance decay problem may be eliminated by making the inactive sections of the display grey instead of black, but, in one experiment, the outline of an inactive section of a direct-view picture tube was visible after 5,000 hours, even though that section had been excited with a 50% grey signal. Techniques have been developed for excitation of the unused areas with signals that vary to match average picture level, however, and those techniques appear to eliminate image striping.[9] No investigation of viewer acceptance of display stripes with varying brightness was found in the research for this paper.

The differential phosphor luminance decay issue is related only to displays using phosphors, such as those based on typical direct-view or projection cathode-ray tubes. Some video projectors, such as the Schlieren-optics-based Eidophor,[10] have never used phosphors, and advanced television displays may be able to take advantage of other phosphor-free technologies.[11,12]

There are two other techniques associated with aspect ratio accommodation, but they require that either the image or the display be effectively non-fixed in shape. One of these techniques is sometimes used in video walls. As shown in Figure 2, a video wall comprised of 4:3 image modules can create a 4:3 image when stacked in a 3 x 3 or 4 x 4 module configuration, but the same modules can create a 16:9 (1.78:1) image when stacked in a 4 x 3 configuration.

When the goal has been not aspect ratio accommodation but the creation of a different aspect ratio than is commonly used in a particular medium, similar modular-screen techniques have been used in many film and video projection systems. These range from the 19th-century Cineorama system (using ten interlocked motion-picture film projectors)[13] to the current *Geographica* theater (using three synchronized video sources) at the National Geographic Society's Explorers Hall in Washington, D.C. The original Cinerama widescreen movie process, using three synchronized film projectors, is probably the most famous of these systems.[2] It has been suggested that, at some future date, consumers will be able to avail themselves of low-cost, non-fixed-aspect-ratio displays, but, at the moment, this technique does not appear to be applicable to most homes.

## Shoot & Protect

The other technique of aspect ratio accommodation may, in fact, be the most common used today, but it cannot be used for existing material shot with just one aspect ratio intended. The technique is sometimes called "shoot &

protect." It has been used to accommodate different aspect ratios both for theatrical film projection and for video display.[14]

In such a system, during production the captured image is framed so as to make images appear in a desired fashion in one aspect ratio while additional area is suitably protected in the overall frame to allow the images to be seen in a different aspect ratio without lighting instruments, masking, microphones, puppeteers, or the edges of set pieces becoming visible. Such framing is facilitated by the presence of reference lines ("reticles") on the camera viewfinder for the desired ("action") aspect ratio.[15] Thus, the inner, action area is sometimes referred to as the reticle region, and the outer frame is sometimes referred to as the aperture.[16] The area between the reticle and the aperture, where significant action is to be avoided, has been referred to as "fluff."

Reconsidering Figure 1-B, in a shoot-and-protect system, the black bands would not be black but would contain, instead, a continuation of the background of the image, the continuation area avoiding anything critical to the action. This is shown in Figure 3.

A shoot & protect system allows aspect ratio accommodation without image truncation (and its associated additional pans or cuts), image shrinking (and its associated reduced viewing angle, reduced resolution, objectionable blank screen bars, and potential differential phosphor decay), and/or image distortion. On the other hand, it creates major restrictions in the way sets can be dressed and lit, the way sound can be picked up, and the way action can be framed. A character cannot be positioned at the edge of a frame, for example, if that edge will not appear in one of the aspect ratios.

This restriction can affect not merely aesthetic shooting preferences but also plot. In the play *Largely New York* (1989), for example, a character trapped in a television signal tries to get out by pushing on the edges of the frame; if that material was captured in a shoot & protect system, in at least one of the aspect ratios those frame edges would not be properly located, so the plot device could not be used. Dramatic and comedic timing is sometimes affected by the moment when a particular character enters a frame; in a shoot & protect system, a character may appear at different times on different displays.

There is an additional problem associated with shoot-and-protect of the form where the action area is the wider aspect ratio (upper portion of Figure 3). This problem relates to theatrical projection framing. With no visual indication of the upper and lower limits of the wider frame, a projectionist must guess at the correct framing, and that framing is not necessarily intended to be centered in the protected aperture.[17,18,19] Despite all of these problems, because home video, alone, has resulted, since 1986, in greater domestic wholesale gross revenues for movie distributors than has theatrical release,[20] there is a strong financial incentive for this form of aspect ratio accommodation, whatever its problems. Aside from its other negative aspects, in 1987, pan & scan (and associated) costs ran as high as $8,000 per feature at one cable television network.[2] There is also a need to consider accommodation of films shot in very wide aspect ratios on much narrower theatrical screens.[17]

It should be pointed out that the viewfinder markings of a shoot & protect system may be used even when there is only one intended aspect ratio, simply to allow the use of imaging equipment designed for a different aspect ratio. The Sony Jumbotron screen at the Skydome in Toronto has an aspect ratio of 10:3 (3.33:1), but its images come from conventional 4:3 television cameras with appropriate viewfinder reticles. In fact, many widescreen films shot this way (with a reticle framing a widescreen image in a conventional 4:3 frame) have been shown on 4:3 screens as though they had been created in a shoot & protect mode.

Unfortunately, such presentation often shows viewers image areas intended by the director and cinematographer not to be seen. It is possible to see a microphone intruding into the top of the image in *Hatari!* (1962), for example, when that movie, intended to be seen on a wider aspect ratio theatrical screen, has its full film frame exhibited on a 4:3 display. Theatrical projection masking would have kept the microphone out of the shot. Nudity intended not to be seen in *Bonnie and Clyde* (1967) is similarly a result of full-frame exhibition of material shot to be shown with much less of the film frame visible. In *Psycho* (1960), set masking is visible when the full 4:3 frame is presented.[2] Such visible microphone booms, masking, set edges, and even lighting instruments have been attributed, in some cases, to sloppy filmmaking; the real cause is exhibition in an aspect ratio never intended by the director or cinematographer.

In addition to the shoot-and-protect systems of Figure 3, it's also possible to create a shoot-and-protect system matching the shape of neither of the aspect ratios needing accommodation but providing both with "equal pain." Such a system would have an outer protection frame (aperture) as high as the narrowest desired aspect ratio and as wide as the widest (when both have equal area) and an inner action frame (reticle) as high as the widest aspect ratio and as wide as the narrowest. The outer and inner frames would have the same aspect ratio. This new shooting aspect ratio is derived from the following formula:

$$\mathbf{D} = \mathbf{W} * (1/\mathbf{W})^{1/2}/(1/\mathbf{N})^{1/2}$$

where **D** is a derived compromise aspect ratio, **W** is the widest of a range of desired aspect ratios, and **N** is the narrowest of the range.

The benefit of such a system would be to reduce the amount of non-action <u>linear</u> dimension in all aspect ratios between the narrowest and widest selected. Since the shoot & protect systems of Figure 3 have <u>no</u> non-action area for one aspect ratio (the reticle), however, that aspect ratio would actually suffer an <u>increase</u> in non-action area through this plan. Of course, if a display screen happened to have the same aspect ratio as the derived shooting system, it could be perfectly framed, with no non-action area.

Figure 4 shows three possibilities for such equal-area-based shoot-and-protect systems. Figure 4-A shows what might be considered a sublime example of such a system. The two extreme aspect ratios being accommodated are so close that the inner action frame is proportionally even larger than that of the SMPTE Safe Action area accommodating overscanning in TV sets.

Figure 4-B shows what might be considered a ridiculous application of such a system. The inner action frame is so small relative to the outer protection frame that it's difficult to see how any director or cinematographer could make meaningful use of such restricted framing.

**16:9**

Figure 4-C shows the actual equal-area-based shoot & protect shape proposed by Kerns Powers, then with RCA Laboratories, at the meeting of the SMPTE Working Group on High Definition Electronic Production on 4 May 1984. It is clearly closer to that of Figure 4-A than to that of Figure 4-B. The aspect ratio was derived from the above formula through the substitution of 4:3 for the narrowest aspect ratio and 2.35:1 (then the projected shape of anamorphically expanded 35 mm theatrical movies, commonly referred to as "scope," from the CinemaScope system) for the widest. The resulting ratio is just over 1.77:1, which was rounded up to 16:9 (somewhat less than 1.78:1) for convenient circuit design.

At the same time, the Motion Picture Association of America was evaluating a new anamorphic projection format using an anamorphic expansion ratio of 1.5:1 instead of the scope 2:1 to improve screen illumination efficiency and reduce projected jitter. That format would have had an aspect ratio of 1.5/2 times the 2.35:1 aspect ratio, or 1.7625:1, very close to the derived shoot & protect 1.77:1.[16]

As a proposed electronic <u>production</u> format, 16:9 need not ever have been used in a shoot & protect mode. Given sufficient resolution, a 16:9 electronic imager could be used to capture single-aspect-ratio moving pictures in any desired aspect ratio. For the extremes of 4:3 and 2.35:1, the 1.77:1 shape would allow minimum waste of photosensitive area (for a multi-aspect-ratio imager) and simplify lens design. In single-aspect-ratio use, the action area could fill the intended display aspect ratio.

Other beneficial properties have been claimed for the 16:9 aspect ratio. The two most common widescreen non-anamorphic theatrical projection aspect ratios worldwide are 1.85:1 and 1.66:1.[19] A linear average of the two is 1.755:1, again close to 1.77:1 (but closer still to the theatrical projection ratio of 1.75:1 adopted by some major film distributors[21] and standardized by 1956[18]). There is also an interesting mathematical progression from 4:3 to 16:9 (4/3 * 4/3) to approximately 2.35:1 (4/3 * 4/3 * 4/3).

It has been reported that the 16:9 aspect ratio was unanimously approved by the Working Group and that a number of cinematographers were involved.[15,16] Such reports have also been disputed.[22] Some of the conflict may simply be semantic (e.g., what makes a person a cinematographer?) For the purposes of this paper, the arguments over who knew or approved what when are not significant. Interested readers are referred to sources listed here and elsewhere in this paper.[23,24]

Even before the Working Group meeting, in 1983 Joseph Nadan, then with Philips Laboratories in the U.S., illustrated possible advantages for a consumer high-definition television (HDTV) set that had a display shape of 16:9. In a "polyscreen" mode, the 16:9 shape could be divided into 12 4:3 images (an inverse of Figure 2's video wall); in MPIP mode, it could provide three stacked 4:3 images adjacent to one larger one[25] (note that some in the consumer electronics industry refer to this mode as "picture-outside-picture," or POP). These display modes are shown in Figure 5.

It was also pointed out that 16:9 was "friendly" to two international recommendations: that HDTV have twice the resolution of non-HD TV and that digital component non-HD TV have 720 active picture elements per scanning line. Twice 720 is 1440 for a 4:3 aspect ratio. For a 16:9 aspect ratio, it would be 1920, and equivalent vertical picture element resolution ("square pixels") would dictate 1080 active scanning lines, for a total of 2,073,600 picture elements (pixels). A two megapixel memory has $2^{21}$ pixels: 2,097,152, a near-perfect match.[26] Even a slight aspect ratio increase to the common widescreen 1.85:1 would require 2,157,840 pixels, a poor fit to common random-access memory sizes.

The 16:9 aspect ratio also allowed for a form of simplified dual aspect ratio transmission. If a 16:9 image is transmitted at the common composite video digital sampling rate of four times the color subcarrier frequency ($4f_{sc}$), a receiver can recover the full 16:9 image by reading its memory at $4f_{sc}$ or can get a 4:3 truncated version (potentially in a pan & scan mode) by reading the memory at $3f_{sc}$.[27] Only aspect ratios that have a 4/3 relationship (as do 16:9 to 4:3 and 2.37:1 to 16:9) can make use of this technique with common sampling rates.

Since "1.85 is far and away the most common aspect ratio for motion pictures filmed in the United States,"[19] the proximity of 16:9 to 1.85:1 (less than 4% difference) could also be considered beneficial for the display of movies. Circuit design generally requires integer values for multipliers and dividers. The 1.85:1 ratio can be expressed as the complex 37:20. The simpler 9:5 ratio is a very close 1.8:1, but it could not make use of the simplified dual-aspect-ratio transmission system described in the last paragraph, nor would it be able to double digital component resolution and still fit in a two megapixel memory. The 16:9 shape is the closest aspect ratio to 1.85:1 offering those other electronic system design benefits.

The 16:9 aspect ratio is also well matched to the technique of economizing by shooting film frames three perforations high instead of the usual four.[28,29,30] Again, it is the 4/3 relationship between the 4:3 aspect ratio and the 16:9 aspect ratio that makes 16:9 an appropriate three-perforation (3-perf) aspect ratio. The same 4/3 relationship also makes possible the modular display shape modification shown in Figure 2.

One more rationale for the selection of 16:9 was probably unknown to proponents of the ratio in the 1980s. In 1930, the standards committee of SMPTE's predecessor, the Society of Motion-Picture Engineers (SMPE), recommended a new method of projecting large screen movies from wider, 50 mm film. The screen shape used for the Society's viewing purposes was 41 feet by 23 feet, 1.78:1 (the Society rounded off its widescreen film aspect ratio recommendation to 1.8:1). The aspect ratio was said to be in line with the desires of the Academy of Motion Picture Arts and Sciences (AMPAS).[31] In 1953, an aspect ratio of approximately 16:9 was again considered as a standard ratio for theatrical projection.[32]

This plethora of beneficial aspects of 16:9 has sometimes been carried too far. It has been claimed, for example, that 16:9 is the only aspect ratio that causes the inner reticles and outer apertures of Figure 4 to have the same shape; a mere glance at Figure 4 indicates that all equal-area shoot & protect aspect ratios have the same property.

The linear position of 16:9 between 1.66:1 and 1.85:1 is also of questionable benefit. A linear average of the extreme aspect ratios of 4:3 and 2.35:1 is just over 1.84:1, a near perfect match for the existing theatrical widescreen aspect ratio of 1.85:1 (though it may be argued that, to obtain the benefits of a 4/3 relationship with 4:3, 16:9, less than 4% smaller, might have been chosen even if the desired aspect ratio were 1.85:1).

The polyscreen and MPIP advantages of 16:9 may also have been overemphasized. While it's true that only 16:9 yields a polyscreen of 12 4:3 images and an MPIP of 3, Figure 6 indicates some polyscreen and MPIP possibilities of the 2:1 and 5:3 aspect ratios, and Table 1 shows that there are many other possibilities.

**Table 1 - Polyscreen and MPIP Possibilities**

| Aspect Ratio | :9 | :1 | Polyscreen | MPIP |
|---|---|---|---|---|
| 8:3 | 24 | 2.67 | 2 | 1 |
| 2:1 | 18 | 2 | 6 | 2 |
| 16:9 | 16 | 1.78 | 12 | 3 |
| 5:3 | 15 | 1.67 | 20 | 4 |
| 8:5 | 14.4 | 1.6 | 30 | 5 |
| 14:9 | 14 | 1.56 | 42 | 6 |
| 32:21 | 13.7 | 1.52 | 56 | 7 |
| 3:2 | 13.5 | 1.5 | 72 | 8 |

Still more combinations are possible if there are multiple columns of MPIP or if the images are not contiguous on the screen. Though it has a 16:9 screen, for example, a recent RCA television receiver allows up to five MPIP images, not merely three.[33]

The display-memory-size benefits of 16:9 HDTV are also predicated on the very specific requirement of doubling the 720 active horizontal pixels of ITU-R Recommendation 601 for a widescreen display. If the common practice of using 704 pixels to represent the picture width is considered instead, even a 1.85:1 display can use common memory devices (twice 704 divided by 4/3 is 1056; $1056^2$ times 1.85 is 2,063,002, well within the two megapixel limit of 2,097,152). Furthermore, while the 1920 active pixels per line of some 16:9 HDTV systems have a very simple relationship to the 720 active pixels of Recommendation 601, there is no such simple relationship between 1080 active scanning lines and the active scanning lines of either 525/59.94 or 625/50 television systems. Even if only 480 active lines are considered for 525/59.94 instead of the more traditional 483 or 484, the resulting simple relationship, 9:4, is different from the horizontal relationship. If a relationship with Recommendation 601 is ignored, a 2:1 display offers a perfect match to two megapixel memories (2048 x 1024), albeit with somewhat less vertical resolution than 1080 active lines.

Benefits derived from 3-perf production have also been questioned. The 3-perf format has been said to be potentially more unsteady than 4-perf, to offer poorer audio frequency response, and to require difficult projector conversion.[34] For the moment, it also has additional costs associated with its being a non-standard format.[35]

A rarely discussed issue is associated with the concept of using identical scanning characteristics in both an equal-area shoot & protect production format and a display. In the case of 16:9, for example, the shoot & protect system would allow the extraction of a 2.35:1 image with 25% non-action area at the sides or a 1.33:1 image with 25% non-action area at the top and bottom. A 16:9 display occupying only the action area would be perfectly framed, with zero "fluff." If, however, the camera and display have identical scanning characteristics, the display cannot be perfectly framed in the action reticule; it must occupy the entire aperture and will, therefore, suffer 43% non-action area -- considerably more than either a 4:3 display or a 2.35:1 display.

This problem is not specific to the 16:9 aspect ratio, only to the concept of an equal-area shoot & protect production system sharing scanning characteristics with a display. Even the 4:3 aspect ratio suffers similarly in some film shoot & protect systems.[35]

That shoot & protect/display scanning issue notwithstanding, for any or all of its benefits, real or imagined, and perhaps for other reasons (such as political or economic considerations), within a few years after its 1984 introduction as a <u>production</u> technique, the 16:9 aspect ratio had been adopted for ATV/HDTV <u>display</u> not only in the U.S. (except for some HDTV transmission system proponents) but also in Europe and Japan, where different aspect ratios had originally been proposed.[36,37] By 1994, 16:9 appeared to be universally accepted, not only for HDTV but also for widescreen video of ordinary resolution in such systems as the European PALplus and Japanese Clearvision.

## Questioning 16:9

In April 1994, however, the American Society of Cinematographers (ASC) presented positions on aspect ratio of advanced television (ATV) distribution and displays at the Artists Rights Symposium, portions of which are excerpted here:

"While the ASC would prefer an aspect that matches our current widest screen production standard of 2.40:1, we realize that practical engineering and manufacturing requirements must also be considered. Thus, the ASC advocates 2:1 as an adequate, if not ideal, standard ratio." [Author's note: the 2:1 expansion of the current 35mm anamorphic projection aperture yields a 2.4:1 image rather than 2.35:1, and the proposed anamorphic theatrical projection system with a 1.5:1 squeeze/expansion, therefore, now yields a 1.8:1 image rather than 1.76:1[38]].

"Every original film work would be mastered and distributed over US ATV (Advanced Television) in its native aspect ratio. This might be 2.4:1, 2.2:1, 1.85:1, 1.66:1 or 1.33:1 (or 1.78:1 if in the current HDTV format).

"The ATV system should be deployed so that all ATV receivers have a 2.0:1 aspect ratio, at any and all standards at which they might operate."[39]

It's worth separating this position into the two parts that affect the choice of a 16:9 aspect ratio. First, the ASC would like all "films" (the term is clearly used loosely, since it is meant to include HDTV productions) mastered and distributed in the aspect ratio for which they were shot. Second they would like ATV receivers to have a 2:1 aspect ratio display.

The first position, excluding any economic issues that may be associated with it (from the mass manufacture of single-aspect-ratio electronic imagers or from potentially increased bit rates for constant-vertical-resolution transmission of wider aspect ratios, for example), is one that should clearly be favored by both filmmakers and viewers. Filmmakers have long complained of the need to compromise their theatrical framing to accommodate television.[2,40] Consumers are already being offered choices for aspect ratio accommodation in some displays and videodiscs. The cable television channel American Movie Classics (AMC) currently offers its widescreen movies in repeated showings on the same day, once in a pan & scan format and the following time in letterbox; if ATV receivers offered the capability of locally accommodating different aspect ratios, consumers could opt for their choice at any time.

A digital transmission system, as is supposed to be used for HDTV, lends itself to the encoding of movies in any aspect ratio. A few bits can indicate to receivers what the picture aspect ratio is as well as carrying any pan & scan or other accommodation information that the filmmaker, distributor, and/or programmer choose to offer. The display memory (necessary in the receiver for inverse bit-rate reduction) can be read in whatever manner the viewer desires, subject to the capabilities of the TV set, of course.

## 2:1

The second ASC position, adopting a 2:1 display aspect ratio, is not as obviously beneficial to either filmmakers or viewers. A 2:1 display aspect ratio solves none of the aspect ratio accommodation problems, except that it favors the widest aspect ratios to the detriment of narrower ones.

As was noted previously, a 2:1 aspect ratio either uses picture memory circuitry poorly or must use horizontal resolution lower than twice that of digital component video. For any given screen width or diagonal measurement, a 2:1 aspect ratio will provide a smaller image than will 16:9. Even a screen of equal area will appear shorter. Only a screen of equal height will clearly be bigger, but, if it is direct-view picture-tube-based, it will also be deeper, heavier, and more expensive.

The cost of direct-view picture tube displays is related to a number of factors. The costs of phosphor screens and shadow masks are related to their area. The cost of electron beam deflection circuitry is related to display width. The cost of a glass bulb is related to its volume, a specification derived from screen area times depth, with depth based on deflection, which is width-related. The overall display cost, therefore, is based somewhere between area and width, or, roughly (for aspect ratios greater than 1:1), on diagonal measurement.

For any given diagonal measurement, the largest possible rectangular display will be square. Narrower aspect ratio displays (of 1:1 or greater) will be larger in area than wider ones. Table 2 shows the combined effects of a fixed diagonal-based screen size and "letterbox" image display on overall image size for five common motion-picture image shapes (expressed as ratios to one, as is common currently in cinematography) and five common existing or proposed display aspect ratios (expressed as integer ratios as is common in display engineering). The data have been normalized so that a 1.33:1 image on a 4:3 display is considered 100%.

**Table 2 - Image Sizes For Letterboxed Equal-Diagonal Displays**

| Display | | | | | |
|---|---|---|---|---|---|
| 4:3 | 100% | 80% | 72% | 61% | 56% |
| 3:2 | 85% | 87% | 78% | 66% | 60% |
| 5:3 | 73% | 92% | 83% | 70% | 64% |
| 16:9 | 67% | 83% | 86% | 72% | 66% |
| 2:1 | 55% | 69% | 77% | 76% | 69% |
| | 1.33:1 | 1.66:1 | 1.85:1 | 2.2:1 | 2.4:1 |

**Image**

While the 2:1 display provides the largest 2.4:1 and 2.2:1 images, it provides the smallest 1.33:1 images. The center column, representing the most common U.S. theatrical film aspect ratio of 1.85:1, indicates that not only does a 16:9 display provide a much larger image than would a 2:1 display but even a 3:2 display provides a larger letterboxed 1.85:1 image than would a 2:1 display. As it was intended to do, the 16:9 ratio provides approximately equal sized images for both extremes of image shape.

A non-picture-tube-based display technology might not have a diagonal-size-related cost basis, however. Tables 3 & 4 compare the same image and display shapes for screens of any size.

Table 3 shows the amount of screen area left blank to fit images into displays of a different aspect ratio. The same figures indicate the amount of resolution reduction from the maximum that the display can deliver. The shaded area indicates reduction of vertical resolution (traditional letterbox); the unshaded area indicates reduction of horizontal resolution (blank screen areas at the sides of the image).

**Table 3 - Blank Screen Area and Screen-based Resolution Reduction for Letterboxed Displays**

| Display | | | | | |
|---|---|---|---|---|---|
| 4:3 | 0% | [illegible] | [illegible] | [illegible] | [illegible] |
| 3:2 | 11% | [illegible] | [illegible] | [illegible] | [illegible] |
| 5:3 | 20% | 0% | [illegible] | [illegible] | [illegible] |
| 16:9 | 25% | 7% | [illegible] | [illegible] | [illegible] |
| 2:1 | 34% | 17% | 8% | [illegible] | [illegible] |
| | 1.33:1 | 1.66:1 | 1.85:1 | 2.2:1 | 2.4:1 |

**Image**

Again, the 2:1 display offers the best results for the widest aspect ratio images and the worst results for the narrowest. Again, for 1.85:1 images, the 16:9 display is superior to all others. Again, the 16:9 display provides roughly equivalent results for the extremes of image shape.

Table 4 shows display resolution reduction caused by letterbox and fixed display memory size. As in Table 3, screens may be of any size. For a fixed memory size, a narrower aspect ratio offers more vertical resolution and a wider aspect ratio offers more horizontal resolution. The data have been normalized so that the narrowest image (1.33:1) on the narrowest display (4:3) shows zero vertical resolution reduction and the widest image (2.4:1) on the widest display (2:1) shows zero horizontal resolution reduction.

**Table 4 - Fixed Memory Size Resolution Reduction for Letterboxed Displays**

| Vertical Resolution Reduction | | | | | Display | Horizontal Resolution Reduction | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0% | 20% | 28% | 39% | 44% | 4:3 | 18% | 18% | 18% | 18% | 18% |
| 6% | 15% | 24% | 36% | 41% | 3:2 | 23% | 13% | 13% | 13% | 13% |
| 11% | 11% | 19% | 32% | 38% | 5:3 | 27% | 9% | 9% | 9% | 9% |
| 13% | 13% | 17% | 30% | 36% | 16:9 | 29% | 12% | 6% | 6% | 6% |
| 18% | 18% | 18% | 26% | 32% | 2:1 | 34% | 17% | 8% | 0% | 0% |
| 1.33:1 | 1.66:1 | 1.85:1 | 2.2:1 | 2.4:1 | | 1.33:1 | 1.66:1 | 1.85:1 | 2.2:1 | 2.4:1 |
| Image | | | | | | Image | | | | |

Again, the 2:1 display offers the best results for 2.2:1 and 2.4:1 images and the worst for 1.33:1 images. Again, the 16:9 display offers the best results for 1.85:1 images.

Tables 2, 3, & 4 all show, as probably seems obvious, a perfect fit for 1.33:1 images on 4:3 display screens. There is a very large installed base of 4:3 displays in TV sets and computer monitors worldwide as well as very large libraries of roughly 1.33:1 moving image programming -- virtually all pre-1953 movies, virtually all television programming to date, virtually all non-theatrical films, and many post-1953 movies. In 1958, 750 pre-1948 movies were sold by Paramount to MCA for $50 million; in 1994, another Paramount transfer, this time of 4,000 4:3 TV episodes (as well as feature films and other desirable properties) was valued at $9.6 billion.[20]

If ATV brings a wider aspect ratio into the home, it may be expected that sports coverage, shopping programs, game and talk shows, other entertainment programming, and even news coverage will eventually migrate to the wider aspect ratio. The older movies and television programs that form the basis of much of the programming of such channels as American Movie Classics, Nick-at-Nite, and Turner Classic Movies will remain 1.33:1, however.

Table 2 also shows that, of common existing or proposed display shapes, 4:3 offers the largest screen for a given cost for technologies with cost related to diagonal measurement, such as direct-view picture tubes. A transition of television to any widescreen aspect ratio will introduce problems relative to the existing 4:3 display and 1.33:1 programming bases.[4] Thus, it has been argued that a 4:3 aspect ratio should be retained for ATV displays.[41]

Unfortunately, as Tables 2, 3, & 4 show, a 4:3 display offers the smallest images, the lowest resolution, and the greatest blank screen area when displaying the widest aspect ratio imagery. Since it is unlikely that homes will have different ATV displays optimized for different aspect ratios of programming, it seems that a compromise display aspect ratio may be desirable.

Using the same formula from which the 16:9 compromise aspect ratio was derived, 2:1 may be seen to be just under the ideal equal-area compromise aspect ratio for the extremes of 1.85:1 on the narrow end and 2.2:1 on the wide end. Those are the widest commonly projected non-anamorphic 35 mm theatrical aspect ratio and the normal 70 mm theatrical projection aspect ratio, respectively. While more favorable to 2.2:1 and 2.4:1 image aspect ratios than is 16:9, 2:1 is less favorable to the most common theatrical 1.85:1 and 1.66:1 aspect ratios[19] and is much less favorable to the 1.33:1 aspect ratio. It would, therefore, be an appropriate compromise display format only if there is some reason to favor the 2.2:1 and 2.4:1 aspect ratios over 1.85:1, 1.66:1, and 1.33:1.

It has been reported in the past that wider aspect ratio films (2.4:1) earn more theatrical revenues than other films.[34] That is definitely not the case at the time of this writing. The highest grossing film of all time, as reported by *Variety* in its February 20-26, 1995 issue, is *E.T. - The Extra Terrestrial* (1982), a movie shot non-anamorphically on 35 mm film and intended for projection at an aspect ratio not exceeding 1.85:1. The second highest grossing film of all time, *Jurassic Park* (1993), was made the same way. The *Variety* list of the highest grossing films of all time may be compared with a listing of their aspect ratios.[42] Such a comparison indicates that while many of the top-100 films were made at a 2.4:1 aspect ratio they account for much less theatrical revenues than do the narrower aspect ratio films on the list. Lists of the top grossing films of 1994 in both domestic and foreign markets in the February 13-19, 1995 issue of *Variety* yield similar results.

While a 2:1 aspect ratio is more favorable to the 2.4:1 theatrical aspect ratio than is 16:9 or any narrower aspect ratio, it does not match it in the same way that a 4:3 display matches 1.33:1 programming. The 2:1 (16:8) aspect ratio is, in fact, considerably closer to 16:9 than it is to 2.4:1. Whatever its disadvantages, a hypothetical 2.4:1 display would at least have the advantage of allowing side panels or drapes to mask unused portions of a screen for all but the few movies that were wider than that aspect ratio; on a 2:1 display, however, side masking fails for the many 2.2:1 or 2.4:1 movies, which would have to be shown in a letterbox format if their aspect ratio was to be preserved.

Part of the dissatisfaction with 16:9 may be related to the fact that the ratio was introduced as a shoot & protect production format, and the concept of shoot & protect involves cropping of non-action area. It has been suggested that the use of 16:9 as a display format precludes the use of letterbox to preserve the composition of material shot in a wider aspect ratio. That is, of course, not true. Any of the aspect ratio accommodation techniques described in this paper can be used on displays of any aspect ratio. A 2:1 display aspect ratio will be no more free of accommodation techniques than is a 16:9 display, and the 2:1 display will have to use those techniques to a greater extent on 1.33:1, 1.66:1, and 1.85:1 programming than will a 16:9 display.

The ASC call for a specific 2:1 display aspect ratio appears to have originated in a presentation by the cinematographer Vittorio Storaro at a formats seminar conducted by the Technology Council of the Motion Picture-Television Industries on 29 January 1994. Seeking standardization on a single aspect ratio, Storaro suggested a linear compromise between HDTV at approximately 1.8:1 and 70 mm theatrical projection at 2.2:1.[43]

Using linear averaging rather than equal-area could be one reason to reject a 16:9 aspect ratio, but, again, a linear average between the same limits of 1.33:1 and 2.35:1 is just over 1.84:1, not 2:1. Another option would be changing the limits. Given the vast libraries of 1.33:1 programming, it seems unreasonable not to consider that ratio, but the upper limit could be extended to the 2.75:1 aspect ratio of such movies as the 1965 epic *The Greatest Story Ever Told* (because such movies were meant to be seen theatrically on curved screens, it is difficult to attribute a particular aspect ratio to them; the chord and the arc of the screen will yield different figures for width[44]). An equal-area compromise between those aspect ratios would be just over 1.91:1; a linear compromise would be just over 2.04:1.

If the few movies with 2.75:1 aspect ratios are to be considered, however, what about those created in the years between the advent of the sound track and the institution of the Academy aperture in 1932, films with an aspect ratio said to be as narrow as 1.15:1?[45,46,47] While a linear compromise between the two new extremes remains near 2:1 (1.95:1), an equal-area compromise reverts, again, to 16:9 (1.78:1).

## History of the Perfect Aspect Ratio

Since accommodation of different aspect ratios necessarily adversely affects the images involved, perhaps it would be worth ignoring issues of compromise between existing aspect ratios and searching for an ideal aspect ratio. In addition to references already listed, there is a wealth of literature on the history of widescreen movies.[48,49,50,51,52,53,54] Most references attribute the impetus behind the current era of widescreen movies to competition with television. In other words, the problem of showing widescreen movies on television was intentional. A study of the literature indicates some other unusual facts:

- Widescreen motion pictures are at least a century old.
- The impetus for many widescreen developments had nothing to do with a preference for a wider aspect ratio.
- The terms *wide film* and *wide screen* have not always indicated a wider aspect ratio.
- Cinematographers and directors have not always favored aspect ratios even as wide as 2:1.

It might be useful to start at the beginning, but it's difficult to say where that beginning is. Motion picture antecedents have been traced to ancient Rome[55,56] and even earlier.[57] Since aspect ratios are as old as rectangular imagery, however, if there is a human preference for a particular aspect ratio, that preference may be considerably older than motion pictures.

A technical paper in 1931 traced an indication of aspect ratio preference to an Egyptian papyrus document dated to 4750 B.C.;[58] that paper was referenced (indirectly) in a debate about the appropriate aspect ratio for advanced television that took place in 1940.[59] Another technical paper, this time referenced directly in the same debate, listed 16 especially "powerful aspect ratios" between 1.236:1 and 3.618:1 in addition to some others that were merely powerful. Both 1.309:1 and 1.809:1 fell into the most powerful category; so did 2.4472:1 and 2.472:1, aspect ratios differing by just one percent, but, according to the paper having no preferred aspect ratio between them.[60]

## The Golden Ratio

One of the most powerful aspect ratios listed was 1.618:1, a rounded version of a mathematical relationship technically called the division in extreme and mean ratio (DEMR) but more commonly referred to as the Golden Section.[61] It is worth noting some of the many names used for this quantity, because they appear frequently in the history of moving picture aspect ratios.

Names for DEMR can be created by combining the adjectives *continuous, divine, golden, medial,* or *sacred* with the nouns *cut, mean, number, proportion, quotient, ratio, rectangle,* or *section.* It has also been called simply *the section,* the *jewel of geometry,* the *middle and two ends,* the *proportional division,* the *whirling squares,* and the more bizarre *he who understands, Faratra, phi*, and *Victoria.* *Dynamic symmetry*, a term that has been incorrectly used to identify DEMR, refers to an aspect ratio of a rectangle that cannot be divided into squares. While a Golden Rectangle meets the criterion of dynamic symmetry, so do rectangles with aspect ratios of the square roots of 2, 3, or 5, all within the range of aspect ratios from 4:3 to 2.35:1. In contrast, 4:3, 3:2, 5:3, 16:9, 1.85:1, 2:1, 2.2:1, and 2.4:1 do not meet the requirements of dynamic symmetry.

The principle of DEMR is quite simple: A line is cut in such a way that the ratio of the whole line to the larger section is the same as that of the larger section to the smaller section. This may be expressed mathematically as

$$x/y = (x+y)/x$$

where **x** is the larger section of a line and **y** is the smaller section. If the whole line is said to have a unit length, then

$$x+y = 1$$

and a quadratic equation is the result:

$$x^2+x-1 = 0$$

If the smaller section becomes the height of a rectangle and the larger the width, the resulting rectangle has an aspect ratio of approximately 1.618:1 (the absolute value of one solution of the equation); if the opposite is done, the resulting rectangle has an aspect ratio of approximately 0.618:1 (the other solution). Both shapes are Golden Rectangles.

The mathematical principle of DEMR has been known for centuries. In the 19th century, Gustav Fechner conducted experiments to find out whether there was a most preferred aspect ratio, and his results seemed to show a preference in the vicinity of the Golden Ratio.[62] A noted physicist and stimulus/sensation investigator, Fechner is considered a pioneer of psychophysics and contributed much to the technologies of both film and video, including the principle that, within certain limits, the intensity of visual stimulation increases as the logarithm of the stimulus (a principle reiterated frequently in the technical literature).[63]

After Fechner's publication of a seemingly preferred aspect ratio, what appeared to be evidence of that ratio was said to be found in works of art dating back to ancient times, and such reports appeared (and continue to appear) in the literature of aesthetics, architecture, art, mathematics, perception, and psychology, e.g., "Much evidence of the conscious use of the proportions of Golden Rectangles can be found in early Greek art and architecture."[64] Even one of the ATV systems proposed to the Federal Communications Commission selected the Golden Section as its aspect ratio.[65]

There appears to be a similarly large body of literature debunking the Golden Ratio as an aesthetic preference, however. One researcher repeated Fechner's experiments and found that the supposed preference appeared to be an artifact of the experimental technique [66] Another found that any shape even vaguely near the Golden Ratio had been considered to be evidence of its use; he nevertheless acknowledged that the raw data "would support hypotheses that suggest that preferred rectangles often have ratios associated with a spread of values containing points from the interval between 0.6 and 0.7" [horizontally oriented ratios between 1.43:1 and 1.67:1].[67]

Indeed, there have been numerous experiments performed with different techniques at different locations, and all seem to show a preference for an aspect ratio in that range for still pictures.[68] For a "Wide Film" symposium conducted by the Technicians, Producers, and Directors branches of AMPAS on 17 September 1930, the Academy's assistant secretary distributed a memorandum stating that "Howell and Dubray, Lane, Westerberg, and Dieterich agree that the most desirable proportions are those approximating 1.618:1, which correspond to those of the so-called 'whirling square' rectangle (also known as the Golden Cut), based on the principles of dynamic symmetry which have predominated in the arts for centuries."

The director Sergei Eisenstein responded in a speech at the meeting that "'Predomination in the arts for centuries' should in itself be a cause for the profoundest suspicion when application is considered to an entirely and basically new form of art, such as the youngest art, the art of cinema." Eisenstein went on to point out that cinema is based on dynamics.[69]

It's easy to see why a dynamic image medium may elicit different aspect ratio preferences from those of a static image medium. A photograph of a skyscraper may be appropriately framed in a vertical image format, while one of a python is more appropriately framed horizontally. In a dynamic medium, however, a horizontal format can tilt down from the tip of the skyscraper to its base; the vertical format can pan the python from tip of tongue to tip of tail. Furthermore, a character may walk into or across a frame or may rise from a chair or descend stairs. It would seem important, therefore, to study aspect ratio preferences specifically for moving image media; unfortunately, it is difficult to find such studies.

### Static vs. Dynamic Image Aspect Ratio Preferences

It has been stated that there is a preference for wider aspect ratios in moving-image media, even if that means sacrificing resolution.[70] A classic case is the Techniscope film format, developed by Technicolor Italiana in 1960, essentially dividing a standard film frame into two much wider aspect ratio frames, thereby losing half the available vertical resolution.[13,45,54] Though used by a number of directors, including Sergio Leone and George Lucas, Techniscope is not commonly used today.

The previously mentioned 1994 Technology Council formats seminar offers anecdotal evidence of aspect ratio preference for moving pictures. As reported in *International Photographer*, "The votes were consistent. The audience always preferred the widest format with the largest image area."[71] The largest image area presented, however, was the 70 mm format at 2.2:1, while the widest was anamorphic 35 mm at 2.4:1. Thus, the largest image area was not the widest, yet, according to that report and others, the largest was preferred.

The test did not include IMAX, with an image area much larger than anything tested but one of the narrowest aspect ratios (1.43:1).[72] From IMAX and other formats, there is anecdotal evidence that viewers may prefer narrower aspect ratios when they are presented on screens very much larger than those of wider aspect ratios.

In a staged event held at Radio City Music Hall in April 1954, Paramount was able to demonstrate its relatively narrower aspect ratio VistaVision format very favorably by comparing it with CinemaScope's wider aspect ratio projected on a smaller area[2] (it's impossible to assign a specific aspect ratio to VistaVision because Paramount allowed "a great deal of latitude with respect to aspect ratio. Our pictures can be played in anything from 4 to 3 up to 2 to 1 in aspect ratio"[73]). Much later, the author's contemporary report of a demonstration of the narrower aspect ratio FuturVision 360 film format at the SMPTE convention on October 28, 1986 stated, "As the FuturVision screen is lowered after the demonstration, the normal, wide theater screen behind it looks as tiny as a television set."[74]

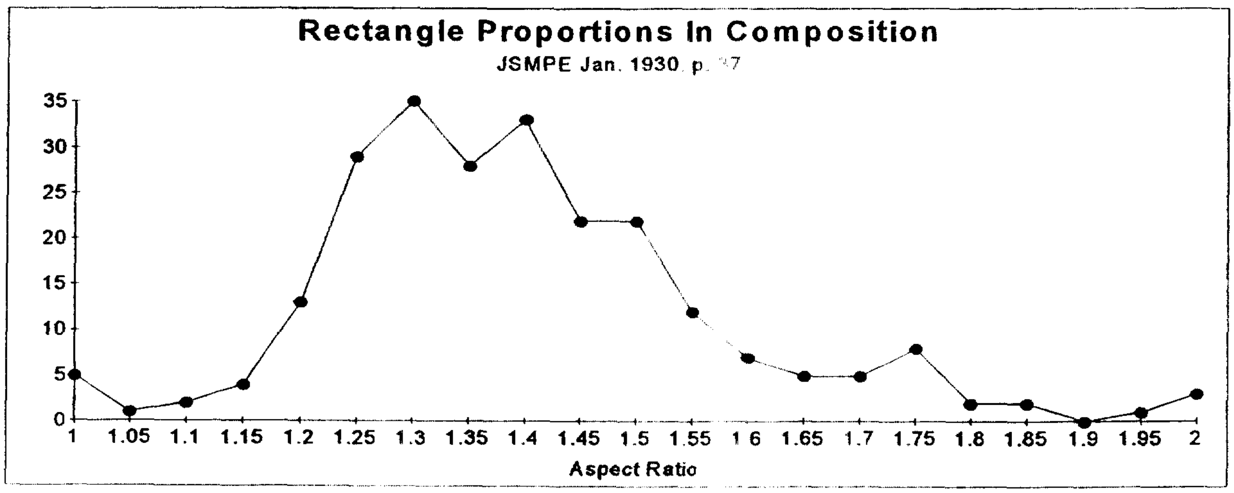
A more formal study found a clear preference for 16:9 moving images over 4:3, even when the 16:9 images are smaller.[75] Unfortunately, only those two aspect ratios were tested. so, while the study may show a preference for widescreen imagery, it doesn't necessarily identify the preferred aspect ratio. It's also possible that the programming selected affected the outcome. The movies chosen were all said to have been selected partly on the basis of their having been shot with both theatrical presentation and television in mind. Thus, shoot & protect was used, with key action likely kept within the confines of the wider aspect ratio. The preference shown for the wider imagery may have been a preference for less fluff in the frame; it's also conceivable that it was a preference for something different from ordinary television.

An unpublished study conducted for Philips using moving images found that aspect ratio viewing preference was influenced slightly by the originally intended aspect ratio. It was also influenced slightly by viewer habit (TV viewers who saw few movies preferred narrower aspect ratios; movie-goers who watched little TV preferred wider aspect ratios) and by viewing angle.[76] The previously mentioned study found no relationship between aspect ratio preference and screen size and contradictory preferences based on viewing distance (screen size and viewing distance are the only factors affecting viewing angle).[75] A third study found a correlation between preferred screen sizes and viewing distances but one that contradicts the results of the other studies.[77] The research for this paper found no clear indication of any particular aspect ratio preference for moving images.

The AMPAS meeting of directors, cinematographers, producers, engineers, and technicians in 1930 was held to determine the best action to take on aspect ratio following the introduction of the sound track. The 4:3 35 mm frame, essentially unchanged since its 1889 introduction in the Edison Kinetoscope, was suddenly narrowed by the addition of a sound track. At approximately the same time, numerous widescreen film techniques were being tried. Virtually the entirety of the January 1930 issue of the *Journal of the SMPE* was devoted to the topic of aspect ratio. No one, it seemed, liked the newer, squarer ratio formed by the sound track, and this seemed an opportune time to change it to something even wider than 4:3.

*Communication No. 410* from the Kodak Research Laboratories was reprinted in the *Journal* as "Rectangle Proportions In Pictorial Composition." The paper came up with yet another term for 1.618:1, "the Golden Rule," and it performed statistical analyses on some 250 museum paintings, specifically excluding those with vertically oriented aspect ratios. A frequency curve was plotted, similar to that in Figure 7.

**Figure 7 - Aspect Ratio Frequency in Museum Paintings**




The thrust of the paper was to have provided impetus for a change in motion-picture aspect ratio, but the average of the aspect ratios shown was just over 1.4:1, and by far the greatest frequencies noted were in the range of the pre-sound-track 4:3 aspect ratio.[47] Perhaps curiously, the exact same technique, averaging the aspect ratios of museum artworks, was used by Paramount's Lorenzo del Riccio to justify the creation of a 1.85:1 aspect ratio.[78]

The differences between the two studies may be related to the artworks selected and/or the measurement techniques used. The inclusion of picture frames results in a narrower aspect ratio, as shown in Figure 8.

Another paper in January 1930 *Journal* was from the Bell & Howell Camera Company and suggested three different film widths, all with a 5:3 aspect ratio.[79] That proposal was particularly significant coming from an organization that previously "had an ironclad company policy to refuse to manufacture, modify, or repair any cinemachine not of the standard 35 mm gauge."[80]

That 5:3 ratio was also referred to as the Golden Rule, a fact explained by the Academy's memorandum: "For simplicity, the ratios 5:3 (which equals 1.667:1) or 8:5 (equaling 1.6:1) are generally advocated instead of 1.618:1."[69] Part of the current aspect ratio debate seems to involve nomenclature,[23] so it is worth pointing out that cinematographers (even ASC members) frequently referred to ratios as 5:3 or 8:5 (or 3:5 and 5:8) at the time of the 1930 debates.[81] It is true that a ratio relating to one provides a more immediate sense of the shape of an aspect ratio than does an integer ratio like 4:3, 16:9, or 64:27; there is a small technical difference, however, between 1.33:1 and 4:3 and an even larger difference between 1.66:1 and 5:3. Again, circuit design also commonly requires integer factors for multiplication or division

Perhaps the most urgent paper in the January 1930 *Journal* was from AMPAS. It described a situation in which standardization had broken down, and both theater owners and movie studios were taking matters into their own hands. Nine different projection apertures and 11 different viewfinder reticles were noted to be in use, none matching any standard, and many with different aspect ratios.[82]

The stage seemed to be set for the first major change in motion-picture aspect ratio. Heads and feet were sometimes being chopped off by arbitrary projection apertures that varied as much as 14% from the standard. There were many proponents of an aspect ratio approximating the Golden Section and some for aspect ratios even wider. Eisenstein's call at the Academy symposium for a "Dynamic Square" has been misinterpreted as a lone call for square imagery. Instead, Eisenstein wanted a square frame into which filmmakers could place any chosen aspect ratio, whether vertical or horizontal.[69] D. W. Griffith in the U.S. and Germaine Dulac in France had previously masked images to highlight certain areas, and, well after the Academy symposium, Eisenstein's dream was realized in the Soviet Union's Vario film systems, the most flexible of which, Vario-70, could deal with any aspect ratio from 0.46:1 to 2.35:1. A short film sponsored by the British Film Institute, *The Door in the Wall* (1955), later also made use of varying aspect ratios, both horizontal and vertical, in a single movie.[13,49]

## 4:3

The amazing result of all the aspect ratio discussions circa 1930 was the Academy aperture, standardized by SMPE in 1932 based on the desires of AMPAS. The projection aperture was precisely 11:8 (1.375:1), very close to the pre-sound-track 4:3[83] (it has since changed to 1.37:1[39]). It was not the first time the roughly 1.33:1 aspect ratio would survive a challenge, and it wouldn't be the last. It's difficult to explain why AMPAS and SMPE reverted to 1.375:1, especially since projection focal lengths and apertures had to be changed anyway. It may have been the case that 11:8 was the only shape on which agreement could be reached (both AMPAS and SMPE had previously tentatively decided on 4:3,[84] but SMPE wanted a firm decision from producers[81]), and agreement on something was certainly necessary, as an article by an ASC member in the 1931 *Journal* noted.[85] That article called for a 4:3 aperture, despite its author's previously expressed preference for the Golden Section.[86]

The widescreen historian John Belton suggests that knowledge of the Golden Section helped create the 4:3 motion-picture aspect ratio in the first place. William Dickson, in developing for Thomas Edison the first motion-picture camera to use flexible transparent film, ordered film 1.375 inches wide (almost 35 mm), because that could be obtained by slitting existing photographic film stock down the middle. Running the film vertically and perforating it dictated an image width of one inch. The height of the image was not dictated mechanically, however. Dickson's probable desire for the Golden Section as an aspect ratio in 1889 was tempered by another desire to work in 1/4" picture increments. Therefore, the first movie frame, one inch by 3/4 inch, 4:3, was as close as he could get to 1.618:1.[2,87]

Unfortunately, some parts of the hypothesis seem weak. Dickson did, indeed, report increasing picture size in 1/4-inch increments,[88] but his perforations appeared 64 times per foot, or every 3/16 of an inch, proving that he worked in increments other than 1/4-inch. As far as picture size is concerned, Edison's first patent caveat, submitted in 1888, describes images just 1/32-inch in size; the third caveat specifies 1/8-inch images.[89]

Had Dickson felt strongly enough about the Golden Section, he could easily have masked the height of the image to that ratio. If that would be considered a waste of film, he could have, as has more recently been suggested,[28,29] made images three perforations high instead of four, thus saving a great deal of film. Though it would have required a different design, the film could also have been moved horizontally through the camera aperture (as in Fear's Super Pictures, Glamorama, VistaVision, Technirama, and IMAX, for example)[54] instead of vertically, thereby removing the one-inch width restriction.

Belton suggests that Dickson may have been influenced by Ottomar Anschutz,[2] whose 1887 animated photography display system used large transparencies in a 3:4 aspect ratio (the opposite of 4:3).[13] One might then question why Anschutz selected 3:4. It may have had to do with the shape of his apparatus or with the recurring ancient Pythagorean 3-4-5 right triangle (a loop of flexible material 12 units of length long, with each unit marked, can be used to create a perfect right angle repeatedly, a principle that was used in the construction of the pyramids). The 4:3 aspect ratio was attributed in 1940 to the ancient Greeks.[59]

There was a plethora of different shapes for photography and animation prior to Dickson's 4:3, and there was a similar plethora afterward. An 1899 survey listed 89 different movie projection systems in its "Present-Day Apparatus" section, many with different aspect ratios, then added another 56 announced systems.[90]

Even though they were using 35 mm film, Auguste and Louis Lumiere began with a 5:4 aspect ratio frame.[91] For compatibility with Edison's movies, they later adopted both a 4:3 aspect ratio and the use of four perforations per side per frame. In 1898, however, when they developed the widest-film motion-picture format (75 mm), for special projections at the Paris Exhibition of 1900, they retained the 4:3 aspect ratio, even though the camera, screen, and projector were all unique and needed no compatibility with anything else.[13] Similarly, Max and Emil Skladanowsky, independent of Edison compatibility and seemingly independent of mechanical requirements, adopted a 4:3 aspect ratio for their first Bioskop projection system.[92]

One of the first post-Kinetoscope wide-aspect ratio systems, the Latham Eidoloscope (1895), was developed by Dickson, who is said to have adopted a wider film, frame, and aspect ratio specifically to avoid infringing Edison patents.[2] Dickson may well have sought to avoid infringing aspects of Edison's patent claims, but none of those claims specified any film size or aspect ratio.[93] Other film systems developed at the time -- even those using wider film -- did not always have an aspect ratio wider than 4:3.[13,54,94]

There was a strong impetus for wider film (but not necessarily wider aspect ratios) regardless of patent infringement issues. That impetus was the requirements of projection versus those of the "peep-show" Kinetoscope viewers for which Dickson had first developed the 4:3 35 mm format.

In an era of nitrate film stock, brightness could not be increased indefinitely without danger of fire; a larger frame, therefore, meant a brighter image. A larger frame also offered benefits related to jitter, resolution, lens magnification, camera and projector mechanical design, and, if the theater could accommodate it, even a larger image.[95] Similar benefits remain true today for larger film formats.[71] Another driving factor for wider film width had nothing to do with pictures; the wider the sound track, the higher the sound quality (a position disputed, however, at the 1930 Academy symposium[81]).

The term *wide film* was clearly defined in an Academy publication: "Wide Film has a width greater than the standard 35 mm."[96] By that definition, current scope movies are not wide film.

There are also references to *wide screens* that indicate simply larger images, not necessarily with a wider aspect ratio.[2,48,97] A publication of the National Association of Theatre Owners states, "'Wide-screen' became the industry watchword for an array of filmmaking techniques and projection systems that delivered high, wide, and mighty images that dwarfed the typical 16-foot by 20-foot theatre screen of the day" (emphasis added).[98]

**Why Wide**

There were, however, considerations favoring wider projected aspect ratios. Key among those was the architecture of the auditoriums in which movies were projected, especially the existence of balconies in movie theaters.[52,53,95,99,100] The overhanging balconies limited sightlines from the rear of the auditorium, placing an absolute limit on picture height, as shown in Figure 9. As movie theaters changed from small, single-level nickelodeons to huge, multi-level palaces, the balcony problem became a serious issue. Today, however, balconies are becoming ever more rare, removing perhaps the major reason for the advent of wider aspect ratios. Belton notes the change in subheadings of his last chapter, "The Return of the Nickelodeon," regarding multiplex theater complexes with small auditoriums, and "The Return of the Peepshow," regarding video.[2]

There were also supposed economic considerations pushing a wider aspect ratio. "Though the opinions of cinematographers were not canvassed, art directors favored the wider frame as it meant they did not have to build sets as high, and production managers favored it because it was felt the larger, clearer images would eliminate the need for closeups and the additional time to shoot them."[53] Even today, it has become necessary to point out that aspect ratio does not determine the height of a scene being shot.[19]

Similar arguments have been made about video production in a high-definition, wider aspect-ratio format -- that it will be possible to use fewer cameras and less editing. Like the impetus created by balconies, the impetus created by any real or imagined economic benefits associated with widescreen production has also vanished, as the publicity about 1995's record-cost widescreen (1.85:1) *Waterworld* indicates.

### Visual Aspect Ratio

Periodically, during aspect ratio debates, allusions have been made to the human visual system -- that there is something about it that would favor one aspect ratio over another (separately from any psychological preferences). As is the case with the Golden Section, however, arguments are often made on both sides. One researcher found that the maximum visual field is approximately twice as wide as it is high,[76] while another found it to be only 1.6:1 for the range captured by the eyes individually and 1:1 for both eyes together.[101]

"A widely spread opinion has it that the screen with a horizontal location, with an aspect ratio of approximately 1:2 [2:1] constitutes the optimum psychophysiological condition. Some authors believe that such a screen format best satisfies the requirements of a full field of view for the two stationary eyes. Such a conclusion is incorrect, however, because the field of distinct vision of the eye is equal to only 2 or 3 degrees. It is only within this small angle that the acuity of vision is approximately 50-100%."[102]

The preceding appeared in the *Journal of the SMPTE* in 1969. Earlier, an article in *Film Quarterly* expressed similar views but expanded them to include wider visual fields, all the way out to peripheral vision, and found that even the widest screens stimulate only a tiny portion of the visual field.[103] Another paper published in the *SMPTE Journal* found important contributions to "sensation of reality" from a wide-field display, however, and that paper, in part, forms the basis for the desire for a wider aspect ratio for HDTV.[104]

Whether a sensation of reality is valuable or not (a director/film-system inventor recently suggested that it can actually interfere with traditional fictional filmmaking[105]) and regardless of how we see, the key to arguments about visual field is the fact that aspect ratio has little or no effect on the retinal angle stimulated by an image. The horizontal visual field angle is determined primarily by the display width and the viewer's distance from it (there are also off-axis contributions); the vertical field is determined by the same distance and the height of the screen. The principle is similar to that used to argue that aspect ratio is not a determinant of scene width during shooting.[19]

The BKSTS recommended theatrical seating plan has the front row no closer than twice the screen width and the rear no farther than six widths.[5] That's a much greater range than the difference in aspect ratios between 4:3 and 2.4:1. Wide angular ranges can also be found in a SMPTE theatrical presentation manual,[106] and common theatrical practice exceeds both BKSTS and SMPTE recommendations. Television also offers widely varying visual angles. Though the preceding argument renders the fact irrelevant, it may be pointed out that the largest motion-picture screens have always had aspect ratios less than 1.4:1.[72,107,108]